EXPERIMENTAL MATHEMATICS
2019, VOL. 28, NO. 2, 151–160
https://doi.org/10.1080/10586458.2017.1360809

Primitive Root Bias for Twin Primes

Stephan Ramon Garcia[a], Elvis Kahoro[a], and Florian Luca[b,c,d]

[a]Department of Mathematics, Pomona College, Claremont, CA, USA; [b]School of Mathematics, University of the Witwatersrand, Wits, Johannesburg, South Africa; [c]Max Planck Institute for Mathematics, Bonn, Germany; [d]Department of Mathematics, Faculty of Sciences, University of Ostrava, Ostrava, Czech Republic

ABSTRACT
Numerical evidence suggests that for only about 2% of pairs p, $p+2$ of twin primes, $p+2$ has more primitive roots than does p. If this occurs, we say that p is *exceptional* (there are only two exceptional pairs with $5 \leqslant p \leqslant 10{,}000$). Assuming the Bateman–Horn conjecture, we prove that at least 0.459% of twin prime pairs are exceptional and at least 65.13% are not exceptional. We also conjecture a precise formula for the proportion of exceptional twin primes.

KEYWORDS
prime; twin prime; primitive root; Bateman–Horn conjecture; Twin Prime Conjecture; Brun Sieve

2010 AMS SUBJECT CLASSIFICATION
11A07; 11A41; 11N36; 11N37

1. Introduction

Let n be a positive integer. An integer coprime to n is a *primitive root* modulo n if it generates the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^\times$ of units modulo n. A famous result of Gauss states that n possesses primitive roots if and only if n is 2, 4, an odd prime power, or twice an odd prime power. If a primitive root modulo n exists, then n has precisely $\varphi(\varphi(n))$ of them, in which φ denotes the Euler totient function. If p is prime, then $\varphi(p) = p-1$ and hence p has exactly $\varphi(p-1)$ primitive roots.

If p and $p+2$ are prime, then p and $p+2$ are *twin primes*. The Twin Prime Conjecture asserts that there are infinitely many twin primes. While it remains unproved, recent years have seen an explosion of closely-related work [Castryck et al. 14, Zhang 14, Maynard 15]. Let $\pi_2(x)$ denote the number of primes p at most x for which $p+2$ is prime. The first Hardy–Littlewood conjecture asserts that

$$\pi_2(x) \sim 2C_2 \int_2^x \frac{dt}{(\log t)^2} \qquad (1–1)$$

in which

$$C_2 = \prod_{p \geqslant 3} \frac{p(p-2)}{(p-1)^2} = 0.660161815\ldots. \qquad (1–2)$$

is the *twin primes constant* [Hardy and Littlewood 04]. A simpler expression that is asymptotically equivalent to (1–1) is $2C_2 x/(\log x)^2$.

A casual inspection (see Table 1) suggests that if p and $p+2$ are primes and $p \geqslant 5$, then p has at least as many primitive roots as $p+2$; that is, $\varphi(p-1) \geqslant \varphi(p+1)$. If this occurs, then p is *unexceptional*. The preceding inequality holds for all twin primes p, $p+2$ with $5 \leqslant p \leqslant 10{,}000$, except for the pairs 2381, 2383 and 3851, 3853.

If p, $p+2$ are primes with $p \geqslant 5$ and $\varphi(p-1) < \varphi(p+1)$, then p is *exceptional*. We do not regard $p = 3$ as exceptional for technical reasons. Let $\pi_e(x)$ denote the number of exceptional primes $p \leqslant x$; that is,

$$\pi_e(x) = \#\{p \leqslant x : p \text{ and } p+2 \text{ are prime and } \varphi(p-1) < \varphi(p+1)\}.$$

Computational evidence suggests that approximately 2% of twin primes are exceptional; see Table 2. We make the following conjecture.

Conjecture 1. A positive proportion of the twin primes are exceptional. That is, $\lim_{x\to\infty} \pi_e(x)/\pi_2(x)$ exists and is positive.

We are able to prove Conjecture 1, if we assume the Bateman–Horn conjecture (stated below). Our main theorem is the following.

Theorem 1. *Assume that the Bateman–Horn conjecture holds.*

(a) *The set of twin prime pairs p, $p+2$ for which $\varphi(p-1) < \varphi(p+1)$ has lower density (as a subset of twin primes) at least 0.459%.*
(b) *The set of twin prime pairs p, $p+2$ for which $\varphi(p-1) \geqslant \varphi(p+1)$ has lower density (as a subset of twin primes) at least 65.13%.*

This article has been republished with minor changes. These changes do not impact the academic content of the article.

CONTACT Stephan Ramon Garcia Stephan.Garcia@pomona.edu Department of Mathematics, Pomona College, 610 N. College Ave., Claremont, CA 91711, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uexm.

Table 1. For twin primes p, $p+2$ with $5 \leqslant p \leqslant 2000$, the difference $\delta(p) = \varphi(p-1) - \varphi(p+1)$ is nonnegative. That is, p has at least as many primitive roots as does $p+2$.

p	$\varphi(p-1)$	$\varphi(p+1)$	$\delta(p)$	p	$\varphi(p-1)$	$\varphi(p+1)$	$\delta(p)$
5	2	2	0	821	320	272	48
11	4	4	0	827	348	264	84
17	8	6	2	857	424	240	184
29	12	8	4	881	320	252	68
41	16	12	4	1019	508	256	252
59	28	16	12	1031	408	336	72
71	24	24	0	1049	520	240	280
101	40	32	8	1061	416	348	68
107	52	36	16	1091	432	288	144
137	64	44	20	1151	440	384	56
149	72	40	32	1229	612	320	292
179	88	48	40	1277	560	420	140
191	72	64	8	1289	528	336	192
197	84	60	24	1301	480	360	120
227	112	72	40	1319	658	320	338
239	96	64	32	1427	660	384	276
269	132	72	60	1451	560	440	120
281	96	92	4	1481	576	432	144
311	120	96	24	1487	742	480	262
347	172	112	60	1607	720	528	192
419	180	96	84	1619	808	432	376
431	168	144	24	1667	672	552	120
461	176	120	56	1697	832	564	268
521	192	168	24	1721	672	480	192
569	280	144	136	1787	828	592	236
599	264	160	104	1871	640	576	64
617	240	204	36	1877	792	624	168
641	256	212	44	1931	768	528	240
659	276	160	116	1949	972	480	492
809	400	216	184	1997	996	648	348

Computations suggest that the value of the limit in Conjecture 1 is approximately 2%; see Figures 1 and 2. A value for the limiting ratio is proposed in Section 5.

It is also worth pointing out that this bias is specific to the twin primes since the set of primes p for which $\varphi(p-1) - \varphi(p+1)$ is positive (respectively, negative) has density 50% as a subset of the primes [Garcia and Luca]. That is, if we remove the assumption that $p+2$ is also prime, then the bias completely disappears. Although only tangentially related to the present discussion, it is worth mentioning the exciting preprint [Lemke Oliver and Soundararajan] which concerns a peculiar and unexpected bias in the primes.

2. The Bateman–Horn conjecture

The proof of Theorem 1 is deferred until Section 4. We first require a few words about the Bateman–Horn conjecture. Let $f_1, f_2, \ldots, f_m$ be a collection of distinct irreducible polynomials with positive leading coefficients. An integer n is *prime generating* for this collection if each $f_1(n), f_2(n), \ldots, f_m(n)$ is prime. Let $P(x)$ denote the number of prime-generating integers at most x and suppose that $f = f_1 f_2 \cdots f_m$ does not vanish identically modulo any prime. The *Bateman–Horn conjecture* is

$$P(x) \sim \frac{C}{D} \int_2^x \frac{dt}{(\log t)^m},$$

in which

$$D = \prod_{i=1}^m \deg f_i \quad \text{and} \quad C = \prod_p \frac{1 - N_f(p)/p}{(1 - 1/p)^m},$$

where $N_f(p)$ is the number of solutions to $f(n) \equiv 0 \pmod{p}$ [Bateman and Horn 62].

If $f_1(t) = t$ and $f_2(t) = t+2$, then $f(t) = t(t+2)$, $N_f(2) = 1$, and $N_f(p) = 2$ for $p \geqslant 3$. In this case, Bateman–Horn predicts (1–1), the first Hardy–Littlewood conjecture, which in turn implies the Twin Prime Conjecture.

Although weaker than the Bateman–Horn conjecture, the Brun sieve [Tenenbaum 15, Thm. 3, Section I.4.2] has the undeniable advantage of being proven. It says that there exists a constant B that depends only on m and D such that

$$P(x) \leqslant \frac{BC}{D} \int_2^x \frac{dt}{(\log t)^m} = (1 + o(1)) \frac{BC}{D} \frac{x}{(\log x)^m}$$

for sufficiently large x. In particular,

$$\pi_2(x) \leqslant \frac{Kx}{(\log x)^2}$$

for some constant K and sufficiently large x. The best known K in the estimate above is $K = 4.5$ [Wu 04].

3. An heuristic argument

We give an heuristic argument which suggests that $\varphi(p-1) \geqslant \varphi(p+1)$ for an overwhelming proportion of twin primes p, $p+2$. It also identifies specific conditions under which $\varphi(p-1) < \varphi(p+1)$ might occur. This informal reasoning can be made rigorous under the assumption of the Bateman–Horn conjecture (see Section 4).

Observe that each pair of twin primes, aside from 3,5, is of the form $6n-1, 6n+1$. Thus, if p, $p+2$ are twin primes with $p \geqslant 3$, then $2|(p-1)$ and $6|(p+1)$. We use this in the following lemma to obtain an equivalent characterization of (un)exceptionality.

Lemma 2. *If p and $p+2$ are prime and $p \geqslant 5$, then*

$$\varphi(p-1) \geqslant \varphi(p+1) \quad \iff \quad \frac{\varphi(p-1)}{p-1} \geqslant \frac{\varphi(p+1)}{p+1}. \tag{3–1}$$

Proof. The forward implication is straightforward arithmetic, so we focus on the reverse. If the inequality on the right-hand side of (3–1) holds, then

$$\begin{aligned} 0 &\leqslant p(\varphi(p-1) - \varphi(p+1)) + \varphi(p-1) + \varphi(p+1) \\ &\leqslant p(\varphi(p-1) - \varphi(p+1)) + \tfrac{1}{2}(p-1) + \tfrac{1}{3}(p+1) \\ &< p(\varphi(p-1) - \varphi(p+1)) + \tfrac{5}{6}p \end{aligned}$$

since $2|(p-1)$ and $6|(p+1)$. For the preceding to hold, the integer $\varphi(p-1) - \varphi(p+1)$ must be nonnegative. □

Table 2. The first 100 exceptional p. Here $\delta(p) = \varphi(p-1) - \varphi(p+1)$.

p	$\delta(p)$	$\pi_2(p)$	$\pi_e(p)$	$\pi_e(p)/\pi_2(p)$	p	$\delta(p)$	$\pi_2(p)$	$\pi_e(p)$	$\pi_e(p)/\pi_2(p)$
2381	− 24	71	1	0.0140845	230,861	− 2304	2427	51	0.0210136
3851	− 72	100	2	0.02	232,961	− 1952	2447	52	0.0212505
14,561	− 240	268	3	0.011194	237,161	− 784	2486	53	0.0213194
17,291	− 16	300	4	0.0133333	241,781	− 4232	2517	54	0.0214541
20,021	− 680	342	5	0.0146199	246,611	− 4440	2557	55	0.0215096
20,231	− 192	344	6	0.0174419	251,231	− 768	2598	56	0.021555
26,951	− 576	430	7	0.0162791	259,211	− 1392	2657	57	0.0214528
34,511	− 736	532	8	0.0150376	270,131	− 3256	2755	58	0.0210526
41,231	− 768	602	9	0.0149502	274,121	− 5376	2788	59	0.0211621
47,741	− 1152	672	10	0.014881	275,591	− 1136	2800	60	0.0214286
50,051	− 1728	706	11	0.0155807	278,741	− 6512	2827	61	0.0215776
52,361	− 2088	731	12	0.0164159	282,101	− 7632	2853	62	0.0217315
55,931	− 432	765	13	0.0169935	282,311	− 720	2855	63	0.0220665
57,191	− 912	780	14	0.0179487	298,691	− 3552	2982	64	0.0214621
65,171	− 552	856	15	0.0175234	300,581	− 3420	3000	65	0.0216667
67,211	− 312	876	16	0.0182648	301,841	− 3840	3012	66	0.0219124
67,271	− 96	878	17	0.0193622	312,551	− 4752	3103	67	0.021592
70,841	− 2492	915	18	0.0196721	315,701	− 9228	3130	68	0.0217252
82,811	− 720	1043	19	0.0182167	316,031	− 5376	3132	69	0.0220307
87,011	− 2112	1084	20	0.0184502	322,631	− 7200	3197	70	0.0218955
98,561	− 2132	1207	21	0.0173985	325,781	− 6012	3230	71	0.0219814
101,501	− 228	1235	22	0.0178138	328,511	− 5440	3259	72	0.0220927
101,531	− 240	1236	23	0.0186084	330,821	− 4284	3283	73	0.0222358
108,461	− 312	1302	24	0.0184332	341,321	− 2928	3354	74	0.0220632
117,041	− 4452	1388	25	0.0180115	345,731	− 5088	3388	75	0.022137
119,771	− 912	1420	26	0.0183099	348,461	− 3348	3413	76	0.0222678
126,491	− 1584	1482	27	0.0182186	354,971	− 7920	3459	77	0.0222608
129,221	− 2736	1508	28	0.0185676	356,441	− 4764	3473	78	0.022459
134,681	− 3420	1559	29	0.0186017	357,281	− 6264	3480	79	0.0227011
136,991	− 1568	1586	30	0.0189155	361,901	− 10,232	3520	80	0.0227273
142,871	− 2688	1634	31	0.0189718	362,951	− 4080	3525	81	0.0229787
145,601	− 2448	1653	32	0.0193587	371,141	− 2736	3580	82	0.022905
150,221	− 1688	1703	33	0.0193776	399,491	− 6048	3800	83	0.0218421
156,941	− 2196	1772	34	0.0191874	402,221	− 11,064	3818	84	0.022001
165,551	− 4768	1848	35	0.0189394	404,321	− 1584	3838	85	0.022147
166,601	− 1772	1855	36	0.019407	406,631	− 752	3862	86	0.0222683
167,861	− 3360	1869	37	0.0197967	410,411	− 15,568	3887	87	0.0223823
173,741	− 56	1909	38	0.0199057	413,141	− 3744	3909	88	0.0225122
175,631	− 3232	1924	39	0.0202703	416,501	− 4272	3934	89	0.0226233
188,861	− 2472	2061	40	0.0194081	418,601	− 12,812	3949	90	0.0227906
197,891	− 1392	2139	41	0.0191678	424,271	− 20,448	3996	91	0.0227728
202,931	− 3672	2179	42	0.0192749	427,421	− 1352	4026	92	0.0228515
203,771	− 720	2190	43	0.0196347	438,131	− 4576	4114	93	0.0226057
205,031	− 1136	2204	44	0.0199637	440,441	− 20,088	4120	94	0.0228155
205,661	− 3288	2208	45	0.0203804	448,631	− 13,536	4184	95	0.0227055
206,081	− 468	2211	46	0.0208051	454,721	− 1044	4232	96	0.0226843
219,311	− 3936	2321	47	0.0202499	464,171	− 912	4299	97	0.0225634
222,041	− 1632	2347	48	0.0204516	464,381	− 2148	4302	98	0.0227801
225,611	− 5088	2381	49	0.0205796	465,011	− 9840	4309	99	0.0229752
225,941	− 432	2385	50	0.0209644	470,471	− 24,336	4341	100	0.0230362

In light of (3–1) and the formula (in which q is prime)

$$\frac{\varphi(n)}{n} = \prod_{q|n}\left(1-\frac{1}{q}\right),$$

it follows that p is exceptional if and only if $p+2$ is prime and

$$\frac{1}{2}\prod_{\substack{q|(p-1)\\ q\geqslant 5}}\left(1-\frac{1}{q}\right) < \frac{1}{3}\prod_{\substack{q|(p+1)\\ q\geqslant 5}}\left(1-\frac{1}{q}\right) \qquad (3\text{–}2)$$

because $2|(p-1)$, $3 \nmid (p-1)$ and $6|(p+1)$. The condition (3–2) can occur if $p-1$ is divisible by only small primes. For example, if $5, 7, 11|(p-1)$, then $5, 7, 11 \nmid (p+1)$ and the quantities in (3–2) become

$$\frac{24}{77}\prod_{\substack{q|(p-1)\\ q\geqslant 13}}\left(1-\frac{1}{q}\right) \quad \text{and} \quad \frac{1}{3}\prod_{\substack{q|(p+1)\\ q\geqslant 13}}\left(1-\frac{1}{q}\right).$$

Since

$$\frac{24}{77} \approx 0.3117 < \frac{1}{3} \qquad \text{and} \qquad 2\cdot 5\cdot 7\cdot 11 = 770,$$

one expects (3–2) to hold occasionally if $p = 770n + 1$. Dirichlet's theorem on primes in arithmetic progressions ensures that $p+2 = 770n+3$ is prime $1/\varphi(770) = 1/240 = 0.4167\%$ of the time. Thus, we expect a small proportion of twin prime pairs to satisfy (3–2). For

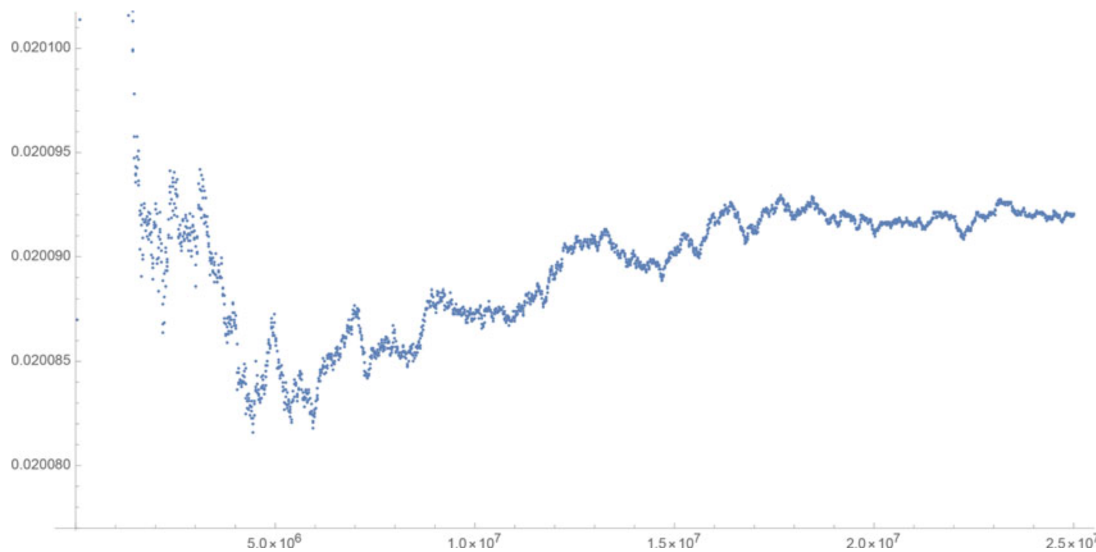

Figure 1. Numerical evidence suggests that $\lim_{x\to\infty}\pi_e(x)/\pi_2(x)$ exists and is slightly larger than 2%. The horizontal axis denotes the number of exceptional twin prime pairs. The vertical axis represents the ratio π_e/π_2.

example, among the first 100 exceptional pairs (see Table 2), the following values of p have the form $770n+1$:

3851, 20021, 26951, 47741, 50051, 52361, 70841, 87011, 98561, 117041, 165551, 167861, 197891, 225611, 237161, 241781, 274121, 278741, 301841, 315701, 322631, 345731, 354971, 357281, 361901, 371141, 410411, 424271, 438131, 440441, 470471.

This accounts for 31% of the first 100 exceptional pairs. We now make this heuristic argument rigorous, under the assumption that the Bateman–Horn conjecture holds.

4. Proof of Theorem 1

Assume that the Bateman–Horn conjecture holds. We first prove statement (a) of Theorem 1. In what follows, p, q, r denote prime numbers.

Proof of (a). Consider twin primes $p, p+2$ such that $5, 7, 11 | (p-1)$. Let $\pi_2'(x)$ be the number of such $p \leqslant x$.

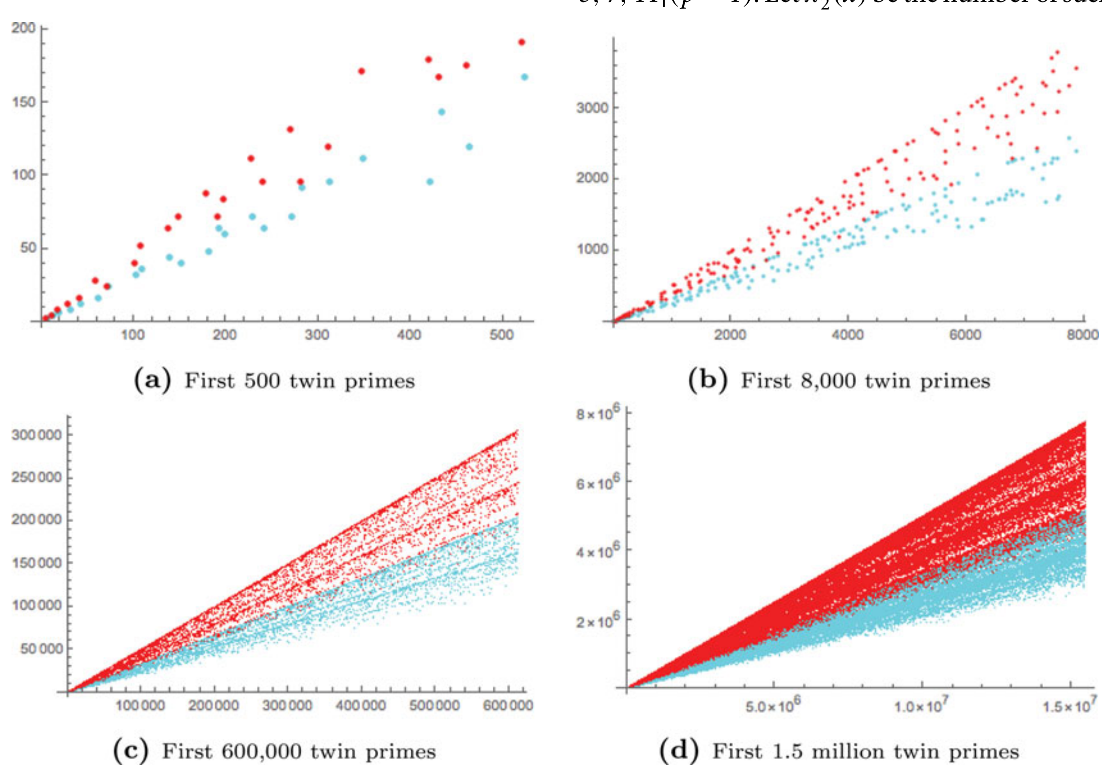

Figure 2. Plots in the xy-plane of ordered pairs $(p, \varphi(p-1))$ (in red) and $(p+2, \varphi(p+1))$ (in cyan) for twin primes $p, p+2$. There are no exceptional pairs visible in Figure 2a; that is, $\varphi(p-1) \geqslant \varphi(p+1)$ in each case. The exceptional pairs 2381, 2383 and 3851, 3853 are visible in Figure 2b. A smattering of exceptional pairs emerge as more twin primes are considered.

Step 1. Since $5\cdot 7\cdot 11 = 385$, the desired primes are precisely those of the form

$$n = 385k+1 \leqslant x \quad \text{such that} \quad n+2 = 385k+3 \text{ is prime.}$$

In the Bateman–Horn conjecture, let

$$f_1(t) = 385t+1, \qquad f_2(t) = 385t+3, \quad \text{and} \quad f = f_1 f_2.$$

Then

$$N_f(p) = \begin{cases} 1 & \text{if } p=2, \\ 2 & \text{if } p=3, \\ 0 & \text{if } p=5,7,11, \\ 2 & \text{if } p\geqslant 13. \end{cases} \tag{4–1}$$

Since $p\leqslant x$, we must have $k \leqslant (x-1)/385$. For sufficiently large x, the Bateman–Horn conjecture predicts that the number of such k is

$$\begin{aligned}
\pi_2'(x) &= (1+o(1))\frac{(x-1)/385}{(\log((x-1)/385))^2}\prod_{p\geqslant 2}\left(\frac{1-N_f(p)/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\left(\frac{2x}{385(\log x)^2}\right)\prod_{p\geqslant 3}\left(\frac{1-N_f(p)/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\left(\frac{2x}{385(\log x)^2}\right) \\
&\quad\times \prod_{p=5,7,11}\left(\frac{1}{(1-1/p)^2}\right)\prod_{\substack{p\geqslant 13\\ \text{or } p=3}}\left(\frac{1-2/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\left(\frac{2x}{385(\log x)^2}\right)\prod_{p\geqslant 3}\left(\frac{1-2/p}{(1-1/p)^2}\right) \\
&\quad\times \prod_{p=5,7,11}(1-2/p)^{-1} \\
&= (1+o(1))\left(\frac{2x}{385(\log x)^2}\right)\prod_{p\geqslant 3} \\
&\quad\times \left(\frac{p(p-2)}{(p-1)^2}\right)\frac{5\cdot 7\cdot 11}{(5-2)(7-2)(11-2)} \\
&= (1+o(1))\frac{2C_2 x}{135(\log x)^2} \\
&= (1+o(1))\frac{\pi_2(x)}{135} \\
&> 0.00740740\,\pi_2(x).
\end{aligned} \tag{4–2}$$

Step 2. Fix a prime $r\geqslant 13$. Let $\pi_{2,r}'(x)$ be the number of primes $p\leqslant x$ such that $p, p+2$ are prime, $5,7,11|(p-1)$, and $r|(p+1)$. The desired primes are precisely those of the form

$$\begin{aligned} &\mathrm{n} = 385k+1 \leqslant x \quad \text{such that} \\ &n+2 = 385k+3 \quad \text{is prime and } r|(385k+2). \end{aligned}$$

In particular, k must be of the form

$$k = k_0 + r\ell,$$

in which k_0 is the smallest positive integer with $k_0 \equiv -2(385)^{-1} \pmod r$. Let $b_r = 385k_0+1$. Then

$$n = 385r\ell + b_r \qquad \text{and} \qquad n+2 = 385r\ell + (b_r+2), \tag{4–3}$$

are both prime, $n\leqslant x$, and

$$\ell \leqslant \frac{x-b_r}{385r}.$$

In the Bateman–Horn conjecture, let

$$f_1(t) = 385rt + b_r, \qquad f_2(t) = 385rt + (b_r+2),$$
and $\quad f = f_1 f_2.$

Then $N_f(p)$ is as in (4–1) except for $p=r$, in which case $N_f(r)=0$. Indeed,

$$\begin{aligned} &f_1(t) \equiv b_r \equiv 385k_0+1 \equiv -1 \pmod r \\ \text{and}\quad &f_2(t) \equiv b_r+2 \equiv 1 \pmod r \end{aligned}$$

for all t. As $x\to\infty$, the Bateman–Horn conjecture predicts that the number of such ℓ is

$$\begin{aligned}
\pi_{2,r}'(x) &= (1+o(1))\frac{(x-b_r)/(385r)}{(\log((x-b_r)/(385r)))^2} \\
&\quad\times \prod_{p\geqslant 2}\left(\frac{1-N_f(p)/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\frac{x}{385r(\log x)^2}\prod_{p\geqslant 2}\left(\frac{1-N_f(p)/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\frac{2x}{385r(\log x)^2}\prod_{p\geqslant 3}\left(\frac{1-N_f(p)/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\frac{2x}{385r(\log x)^2}\prod_{p=5,7,11,r}\left(\frac{1}{(1-1/p)^2}\right) \\
&\quad\times \prod_{p=5,7,11,r}\left(\frac{1-2/p}{(1-1/p)^2}\right) \\
&= (1+o(1))\left(\frac{2x}{385r(\log x)^2}\right)\prod_{p\geqslant 3}\left(\frac{p(p-2)}{(p-1)^2}\right) \\
&\quad\times \frac{5\cdot 7\cdot 11\cdot r}{(5-2)(7-2)(11-2)(r-2)} \\
&= (1+o(1))\frac{2C_2 x}{135(r-2)(\log x)^2} \\
&= (1+o(1))\frac{\pi_2(x)}{135(r-2)}.
\end{aligned} \tag{4–4}$$

Step 3. Suppose that p is counted by $\pi_2'(x)$; that is, suppose that $p, p+2$ are prime and that $5,7,11|(p-1)$. Then $6|(p+1)$, $5,7,11 \nmid (p+1)$, and

$$\frac{\varphi(p-1)}{p-1} \leqslant \prod_{q=2,5,7,11}\left(1-\frac{1}{q}\right) = \frac{24}{77}.$$

If the pair p is unexceptional, then Lemma 2 ensures that

$$\frac{1}{3}\prod_{\substack{r|(p+1)\\ r\geqslant 13}}\left(1-\frac{1}{r}\right)=\frac{\varphi(p+1)}{p+1}\leqslant\frac{\varphi(p-1)}{p-1}\leqslant\frac{24}{77}.$$

Consequently,

$$\prod_{\substack{r|(p+1)\\ r\geqslant 13}}\left(1+\frac{1}{r-1}\right)\geqslant\frac{77}{72},$$

in which r is prime. Let

$$F(p)=\sum_{\substack{r|(p+1)\\ r\geqslant 13}}\log\left(1+\frac{1}{r-1}\right).$$

Step 4. We want to count the twin primes pairs $p, p+2$ with $p\leqslant x$, $F(p)\geqslant\log(77/72)$, and $5, 7, 11|(p-1)$. To do this, we sum up $F(p)$ over all twin primes p counted by $\pi_2'(x)$ and change the order of summation to obtain

$$\begin{aligned}A(x)&=\sum_{\substack{p\text{ counted by}\\ \pi_2'(x)}}F(p)\\ &=\sum_{r\geqslant 13}\pi_{2,r}'(x)\log\left(1+\frac{1}{r-1}\right)\\ &\leqslant\sum_{13\leqslant r\leqslant z}\pi_{2,r}'(x)\log\left(1+\frac{1}{r-1}\right)\\ &\quad+\sum_{z<r\leqslant(\log x)^3}\pi_{2,r}'(x)\log\left(1+\frac{1}{r-1}\right)\\ &\quad+\sum_{(\log x)^3<r\leqslant x}\pi_{2,r}'(x)\log\left(1+\frac{1}{r-1}\right)\\ &=A_1(x)+A_2(x)+A_3(x),\end{aligned}\qquad(4–5)$$

in which z is to be determined later. We bound the three summands separately.

(a) If $13\leqslant r\leqslant z$, then (4–4) asserts that

$$\pi_{2,r}'(x)=(1+o(1))\frac{\pi_2(x)}{135(r-2)}$$

uniformly for $r\in[13,z]$ as $x\to\infty$. For sufficiently large x we have[1]

$$\begin{aligned}A_1(x)&\leqslant(1+o(1))\frac{\pi_2(x)}{135}\left(\sum_{13\leqslant r\leqslant z}\frac{1}{(r-2)}\log\left(1+\frac{1}{r-1}\right)\right)\\ &\leqslant(1+o(1))\frac{\pi_2(x)}{135}\left(\sum_{r\geqslant 13}\frac{1}{(r-2)}\log\left(1+\frac{1}{r-1}\right)\right)\\ &\leqslant(1+o(1))\frac{0.02549678}{135}\pi_2(x)\\ &<0.000188865\,\pi_2(x).\end{aligned}$$

(b) If $z<r\leqslant(\log x)^3$, we use the Brun sieve and manipulations similar to those used to obtain (4–4) to find an absolute constant K such that

$$\pi_{2,r}'(x)\leqslant\frac{K(x/(135r))}{(\log(x/(135r)))^2}$$

for sufficiently large x. Since $r\leqslant(\log x)^3$,

$$\log(x/(135r))\geqslant\log(x^{1/2})\geqslant(\log x)/2$$

holds if $x\geqslant 10^{14}$. Then (1–1) ensures that

$$\pi_{2,r}'(x)\leqslant\frac{4Kx}{135r(\log x)^2}\leqslant\frac{5K\pi_2(x)}{135(r-2)}$$

for sufficiently large x. Now we fix z such that $5K/(135(z-2))<10^{-9}$. Since $\log(1+t)<t$ for $t>0$, for sufficiently large x we obtain

$$\begin{aligned}A_2(x)&=\sum_{z<r\leqslant(\log x)^3}\pi_{2,r}'(x)\log\left(1+\frac{1}{r-1}\right)\\ &\leqslant\frac{5K\pi_2(x)}{135}\sum_{r>z}\frac{1}{r-2}\log\left(1+\frac{1}{r-1}\right)\\ &<\frac{5K\pi_2(x)}{135}\sum_{r>z}\frac{1}{(r-2)(r-1)}\\ &=\frac{5K\pi_2(x)}{135}\sum_{r>z}\left(\frac{1}{r-2}-\frac{1}{r-1}\right)\\ &\leqslant\frac{5K\pi_2(x)}{135(z-2)}\\ &<10^{-9}\,\pi_2(x).\end{aligned}$$

(c) Suppose that $(\log x)^3<r\leqslant x$. By (4–3), the primes counted by $\pi_{2,r}'(x)$ lie in an arithmetic progression modulo $385r$. Thus, their number is at most

$$\pi_{2,r}(x)\leqslant\left\lfloor\frac{x}{385r}\right\rfloor+1\leqslant\frac{x}{385r}+1.$$

Since $\log(1+t)<t$, for sufficiently large x we obtain

$$\begin{aligned}A_3(x)&=\sum_{(\log x)^3<r\leqslant x}\pi_{2,r}'(x)\log\left(1+\frac{1}{r-1}\right)\\ &\leqslant\sum_{(\log x)^3<r\leqslant x}\frac{1}{(r-1)}\left(\frac{x}{385r}+1\right)\\ &\leqslant\frac{x}{385}\sum_{r>(\log x)^3}\frac{1}{r(r-1)}+\sum_{(\log x)^3<r\leqslant x}\frac{1}{r-1}\end{aligned}$$

[1] Since $\log(1+t)\leqslant t$ for $t>0$, the terms of the series are $O(1/r^2)$ and hence it converges rapidly enough for reliable numerical evaluation. `Mathematica` provides the value 0.02549677233701458....

$$
\begin{aligned}
&\leqslant \frac{x}{385}\sum_{r>(\log x)^3}\left(\frac{1}{r-1}-\frac{1}{r}\right)+\int_{(\log x)^3-2}^{x}\frac{dt}{t}\\
&\leqslant \frac{x}{385((\log x)^3-1)}+\left(\log t\Big|_{t=(\log x)^3-2}^{t=x}\right)\\
&\leqslant \frac{2x}{385(\log x)^3}+\log x\\
&=\left(\frac{1}{385C_2\log x}+\frac{(\log x)^3}{2C_2x}\right)\frac{2C_2x}{(\log x)^2}\\
&=(1+o(1))\left(\frac{1}{385C_2\log x}+\frac{(\log x)^3}{2C_2x}\right)\pi_2(x)\\
&< 10^{-9}\,\pi_2(x).
\end{aligned}
$$

Step 5. Returning to (4–5) and using the preceding three estimates, we have

$$
\begin{aligned}
A(x)&=A_1(x)+A_2(x)+A_3(x)\\
&< 0.000188865\,\pi_2(x)+10^{-9}\,\pi_2(x)+10^{-9}\,\pi_2(x)\\
&< 0.000188866\,\pi_2(x).
\end{aligned}
$$

for sufficiently large x.

Step 6. Let $\mathcal{U}(x)$ be the set of primes p counted by $\pi_2'(x)$ that are unexceptional; that is, $\varphi(p-1)/(p-1)\geqslant\varphi(p+1)/(p+1)$ by Lemma 2. As we have seen, if $p\in\mathcal{U}(x)$, then $F(p)\geqslant\log(77/72)$. Thus,

$$
\begin{aligned}
0\leqslant \#\mathcal{U}(x)\log(77/72)&\leqslant\sum_{p\in\mathcal{U}(x)}F(p)\leqslant A(x)\\
&\leqslant 0.000188866\,\pi_2(x),
\end{aligned}
$$

from which we deduce that

$$
\#\mathcal{U}(x)\leqslant\left(\frac{0.000179}{\log(77/72)}\right)\pi_2(x)<0.00281306\,\pi_2(x).
$$

The primes p counted by $\pi_2'(x)$ which are not in $\mathcal{U}(x)$ are exceptional; that is $\varphi(p-1)/(p-1)<\varphi(p+1)/(p+1)$. By (4–2) and the preceding calculation, for large x there are at least

$$
\begin{aligned}
\pi_2'(x)-\#\mathcal{U}(x)&>(0.00740740-0.00281306)\,\pi_2(x)\\
&>0.00459\,\pi_2(x)
\end{aligned}
$$

such primes. This completes the proof of statement (a) from Theorem 1. □

Proof of (b). This is similar to the preceding, although it is much simpler. As before, p, q, r denote primes. If $p, p+2$ are prime and p is exceptional, then

$$
\frac{1}{2}\prod_{\substack{r\mid(p-1)\\ r\geqslant 5}}\left(1-\frac{1}{r}\right)=\frac{\varphi(p-1)}{p-1}\leqslant\frac{\varphi(p+1)}{p+1}\leqslant\frac{1}{3}
$$

since $3\nmid(p-1)$ and $6\mid(p+1)$. If we let

$$
G(p)=\sum_{\substack{r\mid(p-1)\\ r\geqslant 5}}\log\left(1+\frac{1}{r-1}\right),
$$

then $G(p)\geqslant\log(3/2)$ holds for all exceptional primes p. Let $\pi_e(x)$ denote the number of exceptional primes $p\leqslant x$. Then

$$
\begin{aligned}
\pi_e(x)\log(3/2)&\leqslant\sum_{\substack{p\text{ counted}\\ \text{by }\pi_2(x)}}G(p)\\
&=\sum_{\substack{p\text{ counted}\\ \text{by }\pi_2(x)}}\ \sum_{\substack{r\geqslant 5\\ r\mid(p-1)}}\log\left(1+\frac{1}{r-1}\right)\\
&\leqslant\sum_{5\leqslant r\leqslant x}\log\left(1+\frac{1}{r-1}\right)\sum_{\substack{p\text{ counted by }\pi_2(x)\\ p\equiv 1\ (\mathrm{mod}\ r)}}1\\
&\leqslant(1+o(1))\pi_2(x)\sum_{r\geqslant 5}\frac{1}{(r-2)}\log\left(1+\frac{1}{r-1}\right)\\
&< 0.14137\,\pi_2(x),
\end{aligned}
$$

which shows that there are at least

$$
\pi_2(x)-\pi_e(x)\geqslant\pi_2(x)\left(1-\frac{0.14137}{\log(3/2)}\right)>0.6513\,\pi_2(x)
$$

unexceptional primes at most x. □

5. Conjectured density

Below we conjecture a value for the density of the exceptional primes relative to the twin primes. In what follows, we let $P(n)$ denote the largest prime factor of n and let $p(n)$ denote the smallest. We let μ denote the Möbius function and remind the reader that $\mu^2(n)=1$ if and only if $n=1$ or n is the product of distinct primes.

Conjecture 2. The density of the exceptional twin primes is

$$
\lim_{x\to\infty}\frac{\pi_e(x)}{\pi_2(x)}=\lim_{\varepsilon\to 0}\prod_{5\leqslant q\leqslant\frac{1}{\varepsilon}}\left(\frac{q-4}{q-2}\right)\left(\sum_{\substack{a,b\\ \mu^2(ab)=1\\ 5\leqslant p(ab)\leqslant P(ab)\leqslant\frac{1}{\varepsilon}\\ \frac{\varphi(a)}{2a}\leqslant\frac{\varphi(b)}{3b}}}\ \prod_{p\mid ab}\left(\frac{1}{p-4}\right)\right).\tag{5–1}
$$

A few remarks about the imposing expression (5–1) are in order. First of all, for each fixed $\varepsilon>0$, the sum involves only finitely many pairs a, b. Indeed, the condition $\mu^2(ab)=1$ ensures that ab is a product of distinct prime factors. The restriction $5\leqslant p(ab)\leqslant P(ab)\leqslant\frac{1}{\varepsilon}$ implies that only finitely many prime factors are available to form a and b. In principle, the right-hand side of (5–1) can be evaluated to arbitrary accuracy by taking ε sufficiently small. Unfortunately, the number of terms involved in the sum grows rapidly as ε shrinks and we are unable to obtain a reliable numerical estimate from (5–1).

As a brief "sanity check," we also remark that the limit in (5–1), if it exists, is at most 1. Without the condition

$$\frac{\varphi(a)}{2a} \leqslant \frac{\varphi(b)}{3b},$$

the inner sum in (5–1) is

$$\begin{aligned}\sum_{\substack{a,b \\ \mu^2(ab)=1 \\ 5\leqslant p(ab)\leqslant P(ab)\leqslant \frac{1}{\varepsilon}}} \prod_{p|ab}\left(\frac{1}{p-4}\right) &= \sum_{\substack{n \\ \mu^2(n)=1 \\ 5\leqslant p(n)\leqslant P(n)\leqslant \frac{1}{\varepsilon}}} 2^{\omega(n)}\prod_{p|n}\left(\frac{1}{p-4}\right) \\ &= \prod_{5\leqslant p\leqslant \frac{1}{\varepsilon}}\left(1+\frac{2}{p-4}\right) \\ &= \prod_{5\leqslant p\leqslant \frac{1}{\varepsilon}}\left(\frac{p-2}{p-4}\right),\end{aligned}$$

which precisely offsets the first product in (5–1).

To proceed, we need to generalize the functions F and G that appeared in the proof of Theorem 1. Let $\varepsilon > 0$ and define

$$F_\varepsilon(p) = \sum_{\substack{r|(p+1) \\ r\geqslant \frac{1}{\varepsilon}}} \log\left(1+\frac{1}{r-1}\right) \quad \text{and}$$

$$G_\varepsilon(p) = \sum_{\substack{r|(p-1) \\ r\geqslant \frac{1}{\varepsilon}}} \log\left(1+\frac{1}{r-1}\right).$$

Particular instances of these functions have appeared in the proof of Theorem 1 with $\varepsilon = 1/5$ for F_ε (called F) and $\varepsilon = 1/13$ for G_ε (called G), respectively.

Lemma 3. *For $\varepsilon > 0$, the number of twin primes $p \leqslant x$ such that $F_\varepsilon(p) > \varepsilon$ is $O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x))$. The same conclusion holds with F_ε replaced by G_ε.*

Proof. The argument is essentially already in the proof of Theorem 1. We do it only for $F_\varepsilon(p)$ since the argument for $G_\varepsilon(p)$ is similar. We sum $F_\varepsilon(p)$ for $p \leqslant x$ with p, $p+2$ prime and use the fact that $\log(1+t) \leqslant t$ to obtain

$$\begin{aligned}\sum_{\substack{p\leqslant x \\ p,\, p+2 \text{ prime}}} F_\varepsilon(p) &\leqslant \sum_{\substack{p\leqslant x \\ p,\, p+2 \text{ prime}}} \sum_{\substack{q|(p-1) \\ q>\frac{1}{\varepsilon}}} \frac{1}{q-1} \\ &= \sum_{q>\frac{1}{\varepsilon}} \frac{1}{q-1} \sum_{\substack{p,\, p+2 \text{ prime} \\ p\equiv 1 \pmod q}} 1 \\ &= \sum_{q>\frac{1}{\varepsilon}} \frac{\pi_2(x,q,1)}{q-1},\end{aligned}$$

in which $\pi_2(x; q, 1)$ denotes the number of primes $p \leqslant x$ with p, $p+2$ prime and $p \equiv 1 \pmod q$. By the usual argument, the number of twin primes p, $p+2$ with $p \leqslant x$ and $p \equiv 1 \pmod q$ equals the number of $t \leqslant x/q$ such that $qt+1$ and $qt+3$ are prime. The number of them is, by the Brun sieve,

$$\pi_2(x; q, 1) \ll \frac{x}{(q-1)(\log x)^2}.$$

The Prime Number Theorem and Abel summation reveal that

$$\sum_{\substack{p\leqslant x \\ p,\, p+2 \text{ prime}}} F_\varepsilon(p) \ll \frac{x}{(\log x)^2}\sum_{q>\frac{1}{\varepsilon}}\frac{1}{(q-1)^2} \ll \frac{\varepsilon\pi_2(x)}{\log(\frac{1}{\varepsilon})}.$$

If we let

$$\mathcal{A}_\varepsilon = \{p : p,\ p{+}2 \text{ prime and } F_\varepsilon(p) > \varepsilon\},$$

then

$$\#\mathcal{A}_\varepsilon(x)\varepsilon \leqslant \sum_{\substack{p\leqslant x \\ p,\, p+2 \text{ prime}}} F_\varepsilon(p) \ll \varepsilon\left(\log\left(\frac{1}{\varepsilon}\right)\right)^{-1}\pi_2(x),$$

which gives $\#\mathcal{A}_\varepsilon(x) = O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x))$. □

To justify our conjecture, we look at the $\frac{1}{\varepsilon}$-part of p^2-1. We first let $\varepsilon \leqslant 0.5$. We note that $2|(p-1)$, $2|(p+1)$ and $3|(p+1)$ for all twin primes $p \geqslant 5$. For two coprime square-free numbers a, b with $5 \leqslant p(ab) \leqslant P(ab) \leqslant \frac{1}{\varepsilon}$, we say that the twin prime p is of *$\frac{1}{\varepsilon}$-type (a, b)* if

$$p-1 = 2^\alpha \prod_{q|a} q^{\alpha_q} \prod_{q>\frac{1}{\varepsilon}} q^{\gamma_q} \quad \text{and}$$

$$p+1 = 2^\beta 3^\gamma \prod_{q|b} q^{\beta_q} \prod_{q>\frac{1}{\varepsilon}} q^{\delta_q}$$

for some positive α, β, γ, α_q and β_q for $q \mid ab$ and non-negative γ_q, δ_q for $q \geqslant \frac{1}{\varepsilon}$. That is, the prime factors of $p-1$ that are $\leqslant \frac{1}{\varepsilon}$ are exactly the ones dividing $2a$ and the prime factors of $p+1$ that are $\leqslant \frac{1}{\varepsilon}$ are exactly the ones dividing $6b$.

Given ε and (a, b), let

$$c_{a,b} = \prod_{\substack{5\leqslant q\leqslant \frac{1}{\varepsilon} \\ q\nmid ab}} q.$$

Note that

$$\frac{\varphi(p-1)}{p-1} = \frac{1}{2}\frac{\varphi(a)}{a}\prod_{\substack{q|(p-1) \\ q>\frac{1}{\varepsilon}}}\left(1-\frac{1}{q}\right) \quad \text{and}$$

$$\frac{\varphi(p+1)}{p+1} = \frac{1}{3}\frac{\varphi(b)}{b}\prod_{\substack{q|(p+1) \\ q>\frac{1}{\varepsilon}}}\left(1-\frac{1}{q}\right).$$

Since

$$e^{-2y} < 1-y < e^{-y} \qquad \text{for } y < \tfrac{1}{2},$$

it follows that

$$1 - 4\varepsilon < e^{-2\varepsilon} < e^{-F_\varepsilon(p)} = \prod_{\substack{q \mid (p-1) \\ q > \frac{1}{\varepsilon}}} \left(1 - \frac{1}{q}\right)$$

hold for all twin primes $p \leqslant x$ except the ones in $\mathcal{A}_\varepsilon(x)$, a set of cardinality $O((\log(\frac{1}{\varepsilon})^{-1}\pi_2(x))$. Consequently,

$$(1 - 4\varepsilon)\frac{\varphi(a)}{2a} \leqslant \frac{\varphi(p-1)}{p-1}$$

holds for all but $O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x))$ twin primes $p \leqslant x$. Thus, the inequality

$$\frac{\varphi(p-1)}{p-1} \leqslant \frac{\varphi(p+1)}{p+1}$$

implies that

$$\frac{\varphi(a)}{2a} \leqslant (1 - 4\varepsilon)^{-1}\frac{\varphi(b)}{3b}.$$

Let us consider twin primes for which

$$\frac{\varphi(b)}{3b} < \frac{\varphi(a)}{2a} < (1 - 4\varepsilon)^{-1}\frac{\varphi(b)}{3b} \tag{5–2}$$

occurs. Since

$$\frac{\varphi(a)}{2a} = \frac{\varphi(p-1)}{p-1}(1 + O(\varepsilon)) \quad \text{and}$$
$$\frac{\varphi(b)}{3b} = \frac{\varphi(p+1)}{p+1}(1 + O(\varepsilon))$$

for all $p \leqslant x$ with $O((\log(\frac{1}{\varepsilon}))^{-1}\pi_2(x))$ exceptions, it follows that twin primes $p \leqslant x$ for which (5–2) holds have the additional property that

$$\left|\frac{\varphi(p-1)}{p-1} - \frac{\varphi(p+1)}{p+1}\right| = O(\varepsilon). \tag{5–3}$$

Let $\mathcal{B}_\varepsilon$ be the set of twin primes for which (5–3) holds. We make the following additional assumption.

Additional assumption. The number of twin primes $p \leqslant x$ for which (5–3) holds is $O(h(\varepsilon)\pi_2(x))$ for some function $h(y)$ with $h(y) \to 0$ as $y \to 0$.

Assumption (5–3) has been shown to hold when p is only a prime [Garcia and Luca]. That is, the number of primes $p \leqslant x$ such that (5–3) holds is at most $O(h(\varepsilon)\pi(x))$, where $h(\varepsilon)$ tends to zero when $\varepsilon \to 0$. In fact, this was a crucial step in showing that $\varphi(p-1) - \varphi(p+1)$ has no bias if only p is assumed to be prime.

Proving this for primes uses the Turan–Kubilius theorem about the number of prime factors $q \leqslant y$ of $p \pm 1$ when p is prime as the parameter y tends to infinity and also Sperner's theorem from combinatorics. With some nontrivial effort, which involves proving first a Turan–Kubilius estimate for the number of distinct primes $q \leqslant 1/\varepsilon$ of $p-1$ and $p+1$ when p ranges over twin primes up to x, the same program can be applied to prove that the additional assumption holds under the Bateman–Horn conjectures. We do not give further details here.

Assume that the additional assumption holds. Then the set of twin primes $p \leqslant x$ such that

$$\frac{\varphi(p-1)}{p-1} < \frac{\varphi(p+1)}{p+1}$$

is within a set of cardinality $O(h(\varepsilon)\pi_2(x))$ from the set of primes for which

$$\frac{\varphi(a)}{2a} < \frac{\varphi(b)}{3b}. \tag{5–4}$$

With this assumption, we proceed as in [Garcia and Luca, Section 2.11]. Fix $\frac{1}{\varepsilon}$, a, b, and $c = c_{a,b}$. We also fix a residue class for p modulo c which is not $\{0, \pm 1, -2\}$. In this case we need to count natural numbers of the form

$$abct + \kappa,$$

in which κ is fixed such that

- $abct + \kappa \leqslant x$,
- $abct + \kappa$ and $abct + \kappa + 2$ are prime,
- $abct + \kappa - 1$ are divisible by all primes in a and coprime to cb,
- $abct + \kappa + 1$ is divisible by all primes in b (and coprime to ca).

Observe that κ is uniquely determined modulo abc once it is determined modulo c. By the Bateman–Horn conjecture, this number is

$$(1 + o(1))\pi_2(x) \prod_{p \mid abc} \frac{1}{(p-2)}.$$

We next sum this over all $q - 4$ progressions modulo q for which $abct + \kappa$ is not congruent modulo q to some member of $\{0, \pm 1, -2\}$ and for all $q \mid c$ getting an amount of

$$(1 + o(1))\pi_2(x) \prod_{p \mid ab}\left(\frac{1}{p-2}\right)\prod_{p \mid c}\left(\frac{q-4}{q-2}\right)$$
$$= (1 + o(1))\prod_{5 \leqslant q \leqslant \frac{1}{\varepsilon}}\left(\frac{q-4}{q-2}\right)\prod_{q \mid ab}\left(\frac{1}{q-4}\right).$$

We now sum up over all pairs a, b with

$$\frac{\varphi(a)}{2a} < \frac{\varphi(b)}{3b},$$

which yields a proportion of

$$(1 + o(1))\prod_{5 \leqslant q \leqslant \frac{1}{\varepsilon}}\left(\frac{q-4}{q-2}\right)\sum_{\substack{a,b \\ 5 \leqslant p(ab) \leqslant P(ab) \leqslant \frac{1}{\varepsilon} \\ \frac{\varphi(a)}{2a} < \frac{\varphi(b)}{3b}}} \mu^2(ab)\prod_{p \mid ab}\left(\frac{1}{p-4}\right)$$

of $\pi_2(x)$ with a number of exceptions $p \leqslant x$ of counting function $O(h(\varepsilon)\pi_2(x))$. This supports Conjecture 2.

6. Comments

We did not need the full strength of the Bateman–Horn conjecture, just the case $r = 2$ and $D = 1$ for certain specific pairs of linear polynomials $f_1(t)$ and $f_2(t)$. Under this conjecture, we have seen that $\varphi(p-1) \leqslant \varphi(p+1)$ for a substantial majority of twin prime pairs $p, p+2$.

There are a few twin primes $p, p+2$ for which

$$\varphi(p-1) = \varphi(p+1). \tag{6–1}$$

For only such $p \leqslant 100{,}000{,}000$ are

$$5, 11, 71, 2591, 208{,}391, 16{,}692{,}551, 48{,}502{,}931, \\ 92{,}012{,}201, 249{,}206{,}231, 419{,}445{,}251, 496{,}978{,}301.$$

The following result highlights the rarity of these twin primes.

Theorem 4. *The number of primes $p \leqslant x$ with $p+2$ prime and $\varphi(p-1) = \varphi(p+1)$ is $O(x/\exp((\log x)^{1/3})$.*

Proof. Suppose that j and $j+k$ have the same prime factors, let $g = (j, j+k)$, and suppose that

$$\frac{j}{g}r + 1 \qquad \text{and} \qquad \frac{j+k}{g}r + 1 \tag{6–2}$$

are primes that do not divide j. Then

$$n = j\left(\frac{j+k}{g}r + 1\right) \tag{6–3}$$

satisfies $\varphi(n) = \varphi(n+k)$ [Graham et al. 99, Thm. 1]. For k fixed, the number of solutions $n \leqslant x$ to $\varphi(n) = \varphi(n+k)$ which are not of the form (6–3) is less than $x/\exp((\log x)^{1/3})$ for sufficiently large x [Graham et al. 99, Thm. 2].

We are interested in the case $k = 2$ and $n = p-1$, in which $p, p+2$ are prime. If j and $j+2$ have the same prime factors, then they are both powers of 2. Thus, $j = 2$ and $j+k = 4$, so $g = 2$. From (6–2) we see that r is such that

$$r + 1 \qquad \text{and} \qquad 2r + 1$$

are prime. Then $n = 2(2r+1) = p-1$, from which it follows that $p = 4r+3$ and $p+2 = 4r+5$ are prime. Consequently,

$$r + 1, \qquad 2r + 1, \qquad 4r + 3, \qquad \text{and} \qquad 4r + 5,$$

are prime. However, this occurs only for $r = 2$ since otherwise one of the preceding is a multiple of 3 that is larger than 3. □

In particular, the number of primes $p \leqslant x$ for which $p+2$ is prime and $\varphi(p-1) = \varphi(p+1)$ is $o(x/(\log x)^2)$. Assuming the first Hardy–Littlewood conjecture, it follows that the set of such primes has density zero in the twin primes.

Acknowledgments

We thank Tomás Silva for independently computing the ratio $\pi_e(x)/\pi_2(x)$ for large x. We also thank the anonymous referee for suggesting the approach of Section 5.

Funding

SRG supported by NSF grant DMS-1265973, a David L. Hirsch III and Susan H. Hirsch Research Initiation Grant, and the Budapest Semesters in Mathematics (BSM) Director's Mathematician in Residence (DMiR) program.

FL was supported in part by grants CPRR160325161141 and an A-rated researcher award both from the NRF of South Africa and by grant no. 17-02804S of the Czech Granting Agency.

References

[Bateman and Horn 62] P. T. Bateman and R. A. Horn. "A Heuristic Asymptotic Formula Concerning the Distribution of Prime Numbers." *Math. Comput.* 16 (1962), 363–367.

[Castryck et al. 14] Wouter Castryck, Étienne Fouvry, Gergely Harcos, Emmanuel Kowalski, Philippe Michel, Paul Nelson, Eytan Paldi, János Pintz, Andrew V. Sutherland, Terence Tao, and Xiao-Feng Xie. "New Equidistribution Estimates of Zhang Type." *Algebra Number Theory* 8: 9 (2014), 2067–2199.

[Garcia and Luca] S. R. Garcia and F. Luca. "On the Difference in Values of the Euler Totient Function near Prime Arguments." Irregularities in the Distribution of Prime Numbers Research Inspired by Maier's Matrix Method. https://arxiv.org/abs/1706.00392

[Graham et al. 99] S. W. Graham, J. J. Holt, and Carl Pomerance. "On the Solutions to $\varphi(n) = \varphi(n+k)$." In *Number Theory in Progress, Vol. 2 (Zakopane-Kościelisko, 1997)*, edited by K. Győry, H. Iwaniec, and J. Urbanowicz, pp. 867–882. Berlin: de Gruyter, 1999.

[Hardy and Littlewood 04] G. H. Hardy and J. E. Littlewood. "Some Problems of 'Partitio Numerorum'; III: On the Expression of a Number as a Sum of Primes." *Acta Math.* 114: 3 (2004), 215–273.

[Lemke Oliver and Soundararajan] R. Lemke Oliver and K. Soundararajan. "Unexpected Biases in the Distribution of Consecutive Primes." *Proc. Natl. Acad. Sci.* 113: 31, E4446–E4454. (http://arxiv.org/abs/1603.03720).

[Maynard 15] J. Maynard. "Small Gaps between Primes." *Ann. of Math. (2)* 181: 1 (2015), 383–413.

[Tenenbaum 15] G. Tenenbaum. *Introduction to Analytic and Probabilistic Number Theory*, volume 163 of *Graduate Studies in Mathematics*. Third edition. Providence, RI: American Mathematical Society, 2015. Translated from the 2008 French edition by Patrick D. F. Ion.

[Wu 04] J. Wu. "Chen's Double Sieve, Goldbach's Conjecture and the Twin Prime Problem." *Acta Arith.* 114: 3 (2004), 215–273.

[Zhang 14] Y. Zhang. "Bounded Gaps between Primes." *Ann. Math. (2)* 179: 3 (2014), 1121–1174.